Some Statistical Methods and their Application to the Design and Analysis of Experiments in the Biosciences

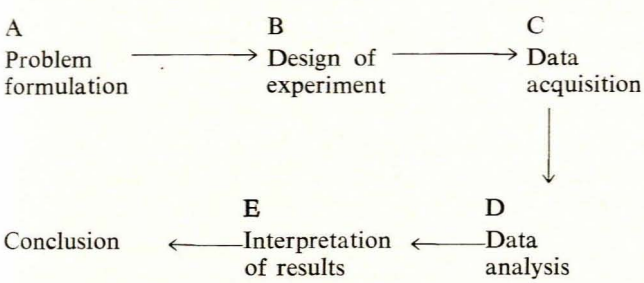
N. F. LAUBSCHER

SUMMARY

The application of statistics to the biological and medical sciences is well known, but has perhaps not yet come into its full right in South Africa. A few examples are described which give an indication of how statistical theory can help the researcher to arrive at statistically valid decisions.

S. Afr. Med. J., **48**, 2525 (1974).

In much of the research work undertaken in any experimental science, the following sequence of procedures has to be adopted before a final conclusion can be arrived at.



Mathematical statistics is the study of properties of mathematical models underlying data analysis. Applied statistics is the science of collecting, analysing and interpreting numerical data relating to an aggregate of individuals. Taking these rather broad definitions into account, it is clear that the statistician has a role to play during the four middle stages of the procedure represented diagrammatically above.

It is essential that any organisation doing experimental work should be able to make use at all times of the services of a statistician. The latter should be a member of the experimental team right from the start. He will at first have a serious problem communicating with the experimenter as his training has been mathematical and he is in most cases unfamiliar with the terminology used by the experimenter. The latter should thus be prepared to spend time explaining all aspects of his proposed project to the statistician; these efforts will be found to pay off handsomely in the long run.

Decisions taken at stage B should be made jointly. The experimenter knows what he requires from his experiment and which plans are feasible. The statistician knows the theory of experimental design and can recommend experimental strategies which will have optimal properties with respect to the data analysis that lies ahead. When stage B has been planned the statistician may very well ask the experimenter to provide him with a set of hypothetical data on which the data analysis techniques to be applied could then be tested. The results so obtained may lead to changes in the experimental design. Computer programmes could also be tested at this stage, and in this way eventual analysis will be greatly facilitated and speeded up. Unfortunately the time factor usually prevents this procedure from being adopted but it is an ideal that should be kept in mind.

As far as stage C is concerned the statistician has at most a supervisory role. He should devote attention to such matters as defining a population frame and he should prescribe methods which will ensure that the samples that are obtained are random.

Stage D is the sole responsibility of the statistician. He will apply methods that will provide answers to the questions put to him. In many cases he also has to help formulate the questions that should be asked by the experimenter!

Stage E is again one of joint decisions. In particular, the statistician should clarify his techniques to the experimenter who will then be in a position to appreciate some of the technical details of the analyses.

This whole discussion serves only to motivate the presence of the statistician at all stages of a project. In what follows I wish to spend some time on some of the details that could go into the data analysis part of the experiment. Naturally I cannot dwell to any extent on the mathematical background of some models usually studied. but I shall attempt to give you an overview of some of the more commonly used methods. Today our subject has become so wide that it is almost impossible for any individual to know even all the textbook material. In most problems current theory cannot be directly applied and in many cases the statistician has to develop his own theory to cater for the intricacies of the specific problem under study. However, I shall deal only superficially with a few of the more popular textbook theories, the use of which have become standard practice today.

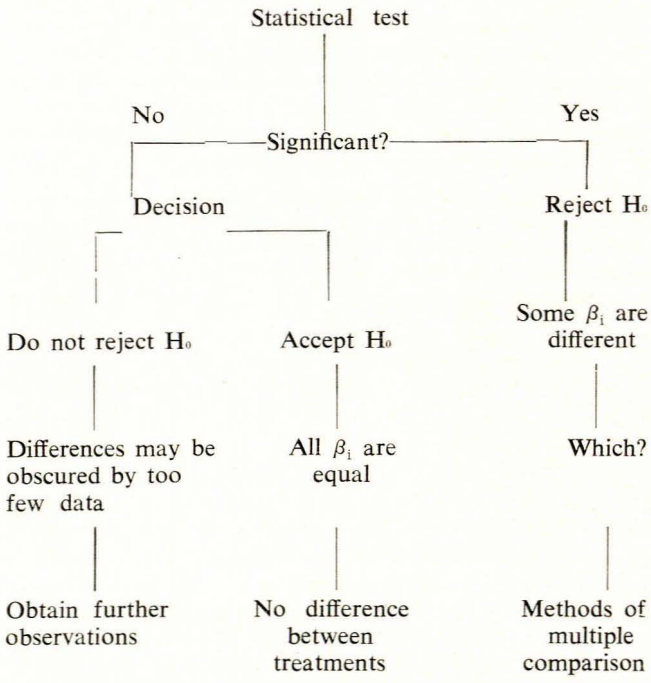
National Research Institute for Mathematical Sciences, CSIR, Pretoria
N. F. LAUBSCHER

Paper presented at the Biennial Congress of the South African Nutrition Society held in Pretoria on 6 - 8 September 1973.

METHODS OF MULTIPLE COMPARISON

This technique is not accepted without controversy even by statisticians, but it has gained favour among the majority. We have found that most experimenters resist these methods, but the reason is probably a lack of appreciation of what is involved. On having the rationale explained to them those not previously exposed to this methodology begin to understand why it is controversial.

Supposing that four randomly selected groups of comparable individuals are subjected to four different treatments. The first might be a control group receiving no treatment at all. The second might be a group receiving the standard treatment, while groups 3 and 4 are subjected to two new treatments. We are measuring some response variable and differences in group means will eventually be ascribed to differences in the treatment. Let the true responses for each group be denoted by unknown quantities β_1, β_2, β_3 and β_4. We wish to test the hypothesis that H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4$, i.e. there is no difference in true responses. Supposing that the assumptions of an analysis of variance F-test are met, then the following series of events and decisions are possible:



If H_0 is rejected it is desirable to make further inferences about the treatment means. The obvious approach is then to make the 6 possible pairwise comparisons by using methods such as the Student t-test. Each of these tests may be done at the α level of significance and we may attach a confidence coefficient of $(1-\alpha)$ to each individual statement that we make. However, this method leaves unanswered the question about the degree of confidence we can have in the correctness of all the statements made. In the above case of 6 statements, each is correct with probability $(1-\alpha)$ hence the probability of r correct statements may be obtained from the binomial distribution and the expected number of correct statements is $6(1-\alpha)$. The probability of all statements being true $(1-\alpha)^6 \approx 0{,}74$ if $\alpha = 0{,}05$ and the probability of at least one false statement is $1-(1-\alpha)^6 \approx 0{,}26$. If the number of statements to be made becomes large the probability of all statements being true approaches zero!

This dilemma is resolved by the method of multiple comparisons which provides an over-all confidence coefficient for the totality of possible statements about tests, the specific ones to be made and the potential ones we do not make. In analysis of variance frameworks there exists, thanks to H. Scheffé, a remarkable procedure for solving the problem. In other models an approximation which is derived from the so-called Bonferroni inequality may be used. In this case, if m statements are to be made, then the over-all probability that all statements are correct will be approximately equal to $(1-\alpha)$ on condition that each individual statement is made with a confidence coefficient of $1-(\alpha/m)$. Thus if we work with $\alpha = 0{,}05$ in the above case of 6 statements, the individual confidence coefficient should be chosen at about 0,992 to guarantee an over-all confidence coefficient of 0,95.

The unacceptability of these procedures stems from the fact that if many statements are to be made, usually only a few individual significant differences can be detected. This unfortunately is the price we have to pay for making correct statements.

EXPERIMENTAL DESIGN

In dealing with this subject I shall again concentrate on only one small aspect; in this case, the determination of sample size.

The size of sample to be used in an experiment cannot be prescribed unless information is available on the variability inherent in the data to be gathered. It is also necessary to specify exactly what is to be done with the data before the sample size can be determined. I shall discuss this problem from the point of view of hypothesis testing. In this connection I shall exploit the concept of the power of the test which is involved.

The significance level of a test guards against rejecting a true null hypothesis. Similarly the power of a test guards against accepting a false null hypothesis. The power of a test is a function of the sample size and of the 'amount' by which the null hypothesis is false.

To fix ideas, it will be assumed we wish to lay out an experiment consisting of a control group of patients receiving no additional vitamin C supplement to a controlled diet. This group has to be compared with two experimental groups, one receiving vitamin C added to the same controlled diet at level I and the other receiving vitamin C added at level II. We wish to test for significant differences (at the 5% level) in serum vitamin C levels of the three groups. How many patients are to be allocated to each group (the same number to each group) so that we can be 90% sure of detecting differences of as little as an amount $\triangle$, if such differences exist?

The point is that the natural variation in the data might obscure differences between groups if an insufficient number of patients is taken. In order to resolve this problem we need to know what kind of variation is to be expected in the data. This information was supplied by the experimenter in the form of the standard deviation based on a previous set of similar data. This standard deviation, together with power curves of the *F*-test, is used to compute the required sample size. The following table (which is purely hypothetical and has nothing to do with real serum vitamin C determinations) is then constructed and presented to the experimenter.

△	0,10	0,15	0,20	0,25
n	182	81	46	30

This leads us to conclude that if about 80 patients are allocated to each group, we would be 90% sure that the *F*-test would eventually declare differences in excess of 0,15 mg/100 ml as significant at the 5% level of significance, if such differences really exist.

MULTIVARIATE ANALYSIS

The world in which we live is essentially multivariate. Many things happen simultaneously and these events are not independent of each other. In the vitamin C example, chances are good that the experimenter would not only be interested in serum vitamin C levels—he is likely to determine quite a number of other biochemical variables, including the cholesterol values. Naturally he would be interested in the interrelationships existing between all the variables measured and in the structure of his data set as a whole. To this effect statisticians have developed a vast theory of multivariate analysis. Typically I could mention multivariate analysis of variance, principal component analysis, canonical analyses, multiple regression analysis and the less statistical but quite useful technique of cluster analysis.

As an example of multivariate analysis I shall talk about the method of discriminatory analysis, but before proceeding, a remark may be in order about the main assumption underlying virtually every existing multivariate technique. This assumption is that the underlying distribution of the variables under study is the multivariate normal. If this assumption is not made the mathematics usually becomes intractable. If a distribution is multivariate normal the marginal distributions (i.e. the distributions of the individual variables) are also single-variate normal. Unfortunately the converse is not always true as was demonstrated in the literature by the exhibition of a (pathological) counterexample. However, if the individual variables have skew distributions, as so many in the biological sciences have, transformations could be applied to make at least the marginals approximately normal. Also, there is a famous result, the so-called central limit theorem, which states that the joint distribution of mean values (when these are based on large samples) is multivariate normal, under some mild additional restrictions. Hence application of normal-theory multivariate analyses may not be too far wrong in practical applications.

Now let us look at the problem of discriminatory analysis. A group of obese but otherwise healthy patients is limited to a mass-reducing diet. In the beginning of the experiment a number of variables are determined on each patient, including a set of biochemical variables obtained from blood analyses and also a set of variables obtained from a battery of psychological tests. At the conclusion of the experiment the amount of reduction in body mass may be used to classify each patient into (say) one of three groups. Let us call these the unsuccessful group (I), an intermediate group (II), and a successful group, (III).

On the basis of the acquired data it may be relevant to be able to predict for a future patient what the probabilities are that he would belong to group I, II or III. If the set of biochemical and psychological variables has any value in discriminating between groups I, II and III, one should be able to do such a classification for individual future patients. Discriminatory analysis was designed for this purpose. A discriminant function, which is a linear combination of all the variables (or of a 'best' subset of these variables) is computed for each of the three groups, and from this the probability of the patient's belonging to each group can then be evaluated.

For the particular example mentioned, the indications are that such a prediction is not expected to be very successful, but, in general, if the discriminatory functions succeed in separating the groups and if it is anticipated that a patient will end in group I, additional therapy might be indicated. If it is predicted that he will fall into group III, the diet alone is indicated and the additional cost of therapy may thus be avoided.

CONCLUSION

I believe that statistical services are indispensable to the experimenter. Unfortunately it costs money to hire a statistician and it costs money to run data on a computer. However, if data analysis is considered an integral part of an experiment, then it seems logical to provide money for that part just as one would have to provide money for buying test tubes.

Then there is a matter which has often in the past been one of embarrassment to me. This is the question of co-authorship when the time of publication comes. The rule should be that if there is a part of the paper which any given individual alone has developed or which he alone can defend scientifically, co-authorship is indicated. The fact that you have paid for the services of a statistician should not influence you in deciding whether to offer him co-authorship. You can never divorce his name from his work.

Finally it remains for me to thank the Council of your Society for inviting me to address you. I greatly appreciate having had this opportunity and hope the occasion will strengthen the existing bonds between two disciplines whose subject matters greatly diverge but which should have quite close ties.